PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A01N 61/00, A61K 45/00, C12Q 1/70, G06T 1/00

A1

(11) International Publication Number: WO 99/17616

(43) International Publication Date: 15 April 1999 (15.04.99)

(21) International Application Number: PCT/US98/20991

(22) International Filing Date: 6 October 1998 (06.10.98)

(30) Priority Data:
60/061,295 7 October 1997 (07.10.97) US

(71) Applicant: NEW ENGLAND MEDICAL CENTER HOSPITALS, INC. [US/US]; 750 Washington Street, Boston, MA 02111 (US).

(72) Inventors: ANDROPHY, Elliot, J.; 2 Mark Street, Natick, MA 01760 (US). BALEJA, James, D.; 99A Appleton Street, Boston, MA 02116 (US). CHEN, Jason, J.; 4068 Park Drive, Boston, MA 02215 (US).

(74) Agents: MYERS, Paul, Louis et al.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: STRUCTURE-BASED RATIONAL DESIGN OF COMPOUNDS TO INHIBIT PAPILLOMAVIRUS INFECTION

(57) Abstract

The invention provides methods of evaluating a compound for the ability to interact with the HPV E6 transforming protein. The invention also provides a machine-readable data storage medium, comprising a data storage material encoded with a set of NMR derived coordinates which define the three-dimensional structure of the E6bp molecule as well as a machine-readable data storage medium, comprising a data storage material encoded with machine readable data which, when used with a machine programmed with instructions for using the data, is capable of displaying a graphical three-dimensional representation of the E6bp molecule. The invention further provides a method of treating a subject at risk for infection by a HPV. A compound having a three-dimensional structure substantially similar to the three-dimensional structure of E6bp and a composition comprising the candidate compound and a carrier macromolecule are also provided. Finally a method of evaluating the ability of a compound to associate with an E6bp molecule is provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

STRUCTURE-BASED RATIONAL DESIGN OF COMPOUNDS TO INHIBIT PAPILLOMAVIRUS INFECTION

FIELD OF THE INVENTION

The invention relates to methods of structure-based rational design of compounds useful for inhibiting infection by papillomavirus.

BACKGROUND OF THE INVENTION

Papillomaviruses (PV) have been linked to widespread, serious human diseases, especially carcinomas of the genital and oral mucosa. It is estimated that there are currently somewhere in the neighborhood of tens of millions of women who suffer from human papilloma virus (HPV) infection of the genital tract. Many of these women eventually develop cancer of the cervix. For example, it has been estimated that about twenty percent (20%) of all cancer related deaths in women worldwide are from cancers which are associated with HPV. It has also been estimated that 90% of all cervical cancer is linked to HPV.

Papillomaviruses induce benign, dysplastic and malignant hyperproliferations of skin or mucosal epithelium (see, for example, Mansur and Androphy, (1993) *Biochim Biophys Acta* 1155:323-345; Pfister (1984) *Rev. Physiol. Biochem. Pharmacol.* 99:111-181; and Broker et al., (1986) *Cancer Cells* 4:17-36, for reviews of the molecular, cellular, and clinical aspects of the papillomaviruses). Almost 70 human papillomavirus types have been identified, and different papillomavirus types are known to cause distinct diseases, Pfister, (1987) *Adv. Cancer Res.*, 48:113-147, Syrjanen, (1984) *Obstet. Gynecol. Survey* 39:252-265. Human papillomaviruses (HPVs) are a heterogeneous group of DNA tumor viruses associated with hyperplastic (warts, condylomata), pre-malignant and malignant lesions (carcinomas) of squamous epithelium. For example, HPV types 1 and 2 cause common warts, and types 6 and 11 cause warts of the external genitalia, anus and cervix. HPV, types 16, 18, 31 and 33 have been isolated from the majority of cervical cancers with HPV-16 present in about 50 percent of all cervical cancers. These HPV's are referred to as "high risk". While HPV 6 and 11 are the most common isolates for cervical warts, these infections rarely progress to invasive cancer, and therefore these HPV's are referred to as "low risk".

Studies of viral gene expression in carcinomas suggest the importance of two HPV encoded proteins, E6 and E7, in malignant development and these proteins have been shown to encode transforming and immortalizing activities. E6-induced

tumorigenesis is thought to occur via multiple pathways (see Turek L., (1994) *Adv. Virus Res.* 44:305; Tommasino M. and Crawford L. (1995) *BioEssays* 17:509; Lee J.M. and Bernstein A. (1995) *Cancer Metas. Rev.* 14:149; Scheffner M. et al., (1993) *Cell* 75:495; and Huibregste J. et al., (1993) *Mol. Cell Biol.* 13:775). In the p53-dependent pathway, the E6 protein associates with a human cellular factor, E6AP and in the p53-independent pathway, the E6 protein associates with a human cellular factor, ERC55.

SUMMARY OF THE INVENTION

In general, the invention features, a method for evaluating a candidate compound for the ability to interact with, e.g., bind, an HPV E6 transforming protein. The method includes: supplying a three-dimensional structure for the E6 binding peptide (E6bp); supplying a three-dimensional structure for the candidate compound; and, optionally, comparing the three-dimensional structure of the candidate compound to the three-dimensional structure of the E6bp, thereby evaluating the candidate compound for the ability to interact with, e.g., bind the HPV E6 transforming protein.

In preferred embodiments, similarity in the structure of the candidate compound to the structure of the E6bp is indicative of the ability of the candidate compound to interact with the HPV E6 transforming protein.

In another aspect, the invention features, a method of providing or identifying a compound, preferably a compound which has the ability to interact with, e.g., bind, an HPV E6 transforming protein. The method includes: supplying a three-dimensional structure for E6bp; supplying a three-dimensional structure for a candidate compound; optionally comparing the three-dimensional structure of the candidate compound to the three-dimensional structure of the E6bp; and optionally altering the structure or altering the spatial position of the structure of the candidate compound, thereby providing or identifying a compound, which preferably has the ability to interact with the HPV E6 transforming protein.

In preferred embodiments, the altered structure of the candidate compound more closely resembles the three-dimensional structure of E6bp, than does the original structure of the candidate compound.

In preferred embodiments, the method includes comparing the altered structure of the candidate compound or the identified compound to the three-dimensional structure of the E6bp. Preferably, the comparison can be performed by defining an atom equivalency in the candidate compound or the identified compound and the E6bp three-dimensional structures and comparing these atom equivalencies.

In preferred embodiments, a second or further subsequent alteration is made in the structure or the spatial position of the structure of the candidate compound.

In preferred embodiments, the method includes defining an atom equivalency in the candidate compound and the E6bp three-dimensional structures; and performing a fitting operation between the candidate compound and the E6bp three-dimensional structures.

In preferred embodiments, the method includes defining an atom equivalency in the candidate compound and the E6bp three-dimensional structures; performing a fitting operation between the candidate compound and the E6bp three-dimensional structures; and analyzing the results of the fitting operation to compare the level of similarity between the candidate compound and the E6bp three-dimensional structures. For example, the atom equivalencies can correspond to protein backbone atoms, e.g., N, Cα, C and O atoms. In preferred embodiments, the fitting operation can be a rigid fitting operation, e.g., the E6bp three-dimensional structure can be kept rigid and the three-dimensional structure of the candidate compound can be translated and rotated to obtain an optimum fit with the rigid target E6bp structure.

In preferred embodiments, the comparison between the candidate compound and the E6bp three-dimensional structures can be performed computationally, e.g., by calculating the root mean square deviation of a set of structural coordinates in the candidate compound from a set of structural coordinates in the E6bp, or visually, e.g., by visual inspection of the candidate compound and the E6bp three-dimensional structures, displayed in a graphical format.

In preferred embodiments, the candidate compound can have an α-helical structure and the alteration can result in a change in the class of the α-helix comprising the structure of the candidate compound. For example, the α-helix comprising the structure of the candidate compound can be selected from the group consisting of A, G, and Y α-helices.

In preferred embodiments, the method includes creating a record of one or more of the three-dimensional structures of the candidate compound, the altered candidate compound, the identified compound, and E6bp. The record can be encoded in the form of a machine-readable data storage medium. The three-dimensional structures can be displayed on a machine capable of displaying a graphical three-dimensional representation.

In preferred embodiments, the method includes providing the identified compound, e.g., chemically synthesizing the identified compound based on the structure identified using the methods described herein. In preferred embodiments, the method

includes assessing the biological activity of the identified compound. The biological activity of the identified compound can be assessed *in vitro*, e.g., in a GST-E6 binding assay or a two-hybrid assay, or *in vivo*, e.g., by applying the compound to a cell line (Hela, Caski, Siha) which expresses HPVE6 and examining the growth characteristics of the cells; or by its tumor suppression ability in an animal model for HPV infection. In preferred embodiments, the identified compound can be combined with a carrier suitable for introduction into an animal model, e.g., naturally derived or synthetic polymers, solvents, dispersion media, coatings, antibacterial and antifungal agents and the like.

In preferred embodiments, the candidate compound can be altered so as to have a three-dimensional structure that is substantially similar to the three-dimensional structure of E6bp provided in figure 1, such that the candidate compound can bind to the HPV E6 transforming protein or portion thereof. For example, the candidate compound can be a peptide, a peptidomimetic, e.g., an isostere, an "inverso" or a "retro-inverso" peptide and the like, or a non peptide organic or inorganic compound.

In preferred embodiments, the identified compound associates with the HPV E6 transforming protein or a portion thereof, such that the ERC55 protein is inhibited from binding to E6. The association may be non-covalent or it may be covalent. The association can be energetically favored by hydrogen bonding or van der Waals or electrostatic interactions.

In preferred embodiments, the three-dimensional structures can be supplied as a set of coordinates, defining the three-dimensional structures of the E6bp molecule, the candidate compound, the altered candidate compound and the identified compound or as a graphical three-dimensional representation of the E6bp molecule, the candidate compound, the altered candidate compound and the identified compound.

In another aspect, the invention features, a machine-readable data storage medium, including a data storage material encoded with a set of NMR derived coordinates which define the three-dimensional structure of the E6bp molecule. The storage medium can be used in methods of the invention.

In yet another aspect, the invention features, a machine-readable data storage medium, including a data storage material encoded with machine readable data which, when used with a machine programmed with instructions for using the data, is capable of displaying a graphical three-dimensional representation of the E6bp molecule. The storage medium can be used in methods of the invention.

In another aspect, the invention features, a method of treating a subject at risk for infection by a HPV. For example, a subject at risk for an HPV induced cancer, e.g., cervical cancer, can be treated. The method includes: administering to a subject a

therapeutically effective amount of a compound other than an E6bp fragment, wherein the compound has a structure sufficiently duplicative of that of Figure 1, so as to bind to the HPV E6 transforming protein and prevent its interaction with the ERC55 protein, thereby treating a subject at risk for infection by a HPV.

In preferred embodiments, 50, 60, 70 and more preferably 80, 90 or 100% of the HPV E6 protein, present in the cell, can be bound to the compound and can, therefore, be unable to bind to ERC55 and induce cellular transformation.

In another aspect, the invention features, a compound, other than an E6bp or ERC55 fragment, having a three-dimensional structure substantially similar to the three-dimensional structure of E6bp provided in Figure 1, such that the candidate compound can bind to the HPV E6 transforming protein.

In preferred embodiments, the candidate compound has a structure sufficiently duplicative of the three-dimensional structure E6bp provided in Figure 1, such that the candidate compound can bind to the HPV E6 transforming protein, with an affinity which is at least half that of E6bp. The dissociation constant (Kd) for the E6-compound complex, is less than 100, 50 or 10 times the Kd of the E6-E6bp complex, and more preferably less than the Kd of the E6-E6bp complex.

In preferred embodiments, the candidate compound is more stable (e.g., more resistant to proteolytic degradation) than E6bp or ERC55.

In another aspect, the invention features, a composition comprising a compound, other than an E6bp or ERC55 fragment, having a three-dimensional structure sufficiently duplicative of the three-dimensional structure of E6bp provided in Figure 1, such that the compound can bind to the HPV E6 transforming protein, and a carrier macromolecule suitable for the administration of the composition to a subject.

In another aspect, the invention features, a method of providing or identifying a compound, preferably a compound which can interact with, e.g., bind E6bp. The method includes: supplying a three-dimensional structure for the E6bp molecule; supplying a three-dimensional structure for the candidate compound; optionally comparing the three-dimensional structure of the candidate compound to the three-dimensional structure of the E6bp; optionally altering the structure or altering the spatial position of the structure of the candidate compound; optionally comparing the altered structure of the candidate compound to the structure of the E6bp, thereby providing or identifying a compound, preferably a compound which can interact with, e.g., bind E6bp.

In preferred embodiments, comparing includes performing a fitting operation.

In another aspect, the invention features, a method of evaluating the ability of a candidate compound to interact with, e.g., bind an E6bp molecule. The method includes: supplying a three-dimensional structure for the E6bp molecule; supplying a three-dimensional structure for the candidate compound; and performing a fitting operation between the three-dimensional structures of the candidate compound and the E6bp molecule, thereby evaluating the ability of the candidate compound to interact with the E6bp molecule.

In preferred embodiments, the method includes analyzing the results of the fitting operation to quantify the association between the candidate compound and the E6bp molecule.

In preferred embodiments, the method includes defining an atom equivalency in the candidate compound and the E6bp molecule three-dimensional structures. For example, the atom equivalencies can correspond to protein backbone atoms, e.g., N, Cα, C and O atoms. In preferred embodiments, the fitting operation can be a rigid fitting operation, e.g., the E6bp three-dimensional structure can be kept rigid and the three-dimensional structure of the candidate compound can be translated and rotated to obtain an optimum fit with the rigid target E6bp structure.

In preferred embodiments, the fitting operation can be performed computationally, e.g., by calculating the root mean square deviation of a set of structural coordinates in the candidate compound from a set of structural coordinates in the E6bp, or visually, e.g., by visual inspection of the candidate compound and the E6bp three-dimensional structures, displayed in a graphical format.

In preferred embodiments, the method includes altering the structure or altering the spatial position of the structure of the candidate compound.

In preferred embodiments, the method includes creating a record of one or more of the three-dimensional structures of the candidate compound, the altered candidate compound, the identified compound and the E6bp molecule. The record can be encoded in the form of a machine-readable data storage medium. The three-dimensional structures can be displayed on a machine capable of displaying a graphical three-dimensional representation.

In preferred embodiments, the method includes providing the identified compound, e.g., chemically synthesizing the identified compound based on the structure identified using the methods described herein. In preferred embodiments, the method includes assessing the biological activity of the identified compound. The biological activity of the identified compound can be assessed *in vitro*, e.g., in a GST-E6bp or a GST-ERC55 binding assay or a two-hybrid assay, or *in vivo*, e.g., by its tumor

suppression ability in an animal model for HPV infection. In preferred embodiments, the identified compound can be combined with a carrier suitable for introduction into an animal model, e.g., naturally derived or synthetic polymers, solvents, dispersion media, coatings, antibacterial and antifungal agents and the like.

In preferred embodiments, the candidate compound can have an α-helical structure and the alteration can result in a change in the class of the α-helix comprising the structure of the candidate compound. For example, the α-helix comprising the structure of the candidate compound can be selected from the group consisting of A, G, and Y α-helices.

In preferred embodiments, the altered compound can associate with the E6bp molecule with a higher affinity.

In preferred embodiments, evaluating includes determining the ability of a compound to interact with, e.g., bind the E6bp molecule. Evaluation can be performed computationally, e.g., by calculating the root mean square deviation of a set of structural coordinates in the candidate compound from a set of structural coordinates in the E6bp, or visually, e.g., by visual inspection of the candidate compound and the E6bp three-dimensional structures, displayed in a graphical format.

In preferred embodiments, the candidate compound includes a compound which can be altered so as to have a three-dimensional structure that is suitable for associating with E6bp and, therefore, with ERC55. For example, the candidate compound can be a peptide, a peptidomimetic, e.g., an isostere, an "inverso" or a "retro-inverso" peptide and the like, or a non peptide organic compound.

In preferred embodiments, the candidate compound associates with the E6bp, such that the ERC55 protein is inhibited from binding to E6. The association may be non-covalent or it may be covalent. The association can be energetically favored by hydrogen bonding or van der Waals or electrostatic interactions.

In preferred embodiments, the three-dimensional structures can be supplied as a set of coordinates, defining the three-dimensional structures of the E6bp molecule, the candidate compound, the altered candidate compound and the identified compound or as a graphical three-dimensional representation of the E6bp molecule, the candidate compound, the altered candidate compound and the identified compound.

In another aspect, the invention features, a method of treating a subject at risk for infection by a HPV. For example, a subject at risk for an HPV induced cancer, e.g., cervical cancer, can be treated. The method includes: administering to a subject a therapeutically effective amount of a compound, wherein the compound associates with E6bp with an affinity so as to prevent the interaction between the ERC55 and the HPV

E6 protein, thereby treating a subject at risk for infection by a HPV. Preferred compounds are provided by the methods described herein.

In another aspect, the invention features, a method of modeling the region of the HPV E6 protein which binds ERC55. The method includes: supplying a three-dimensional structure for an E6bp molecule and supplying a structure, which is complementary to the structure of the E6bp molecule, thereby modeling the region of the HPV E6 protein which binds ERC55.

The molecular modeling techniques, described herein can be used to construct a structure, which is complementary to the E6bp three-dimensional structure. By "complementary", is meant a structure, which is complementary to one or more of: (a) the shape, (b) the electrostatic properties or (c) the hydrophobicity of the E6bp three-dimensional structure. While not wishing to be bound by theory, E6bp may be complementary in shape to a critical portion of E6. Thus, something complementary to E6bp mimics the structure of E6. The complementary structure need not be translated into a real molecule, but can be used in the computational or computer based methods described herein, to identify a compound which has the ability to interact with the HPV E6 transforming protein.

In preferred embodiments, the method includes creating a record of the three-dimensional structure of the E6bp molecule and its complementary structure. The record of the three-dimensional structure of the E6bp molecule and its complementary structure can be encoded in the form of a machine-readable data storage medium. The three-dimensional structures can be displayed on a machine capable of displaying a graphical three-dimensional representation of a structure.

In preferred embodiments, the three-dimensional structure can be supplied as a set of coordinates, defining the three-dimensional structure of the E6bp molecule and its complementary structure or as a graphical three-dimensional representation of the E6bp molecule and its complementary structure.

As used herein, the term "comparing" refers to examining a quality, e.g., three-dimensional structure, hydrophobicity, steric bulk, electrostatic properties, bond angles, size or molecular composition of a compound, in order to identify resemblances or differences between two structures.

As used herein, the term "altering the structure" refers to altering the intrinsic properties, e.g., three-dimensional structure, hydrophobicity, steric bulk, electrostatic properties, bond angles, size or molecular composition of a compound. In a peptide molecule, the alteration can include an amino acid substitution or the introduction of a non-peptide molecule or bond in the structure of the candidate compound. The non-

peptide molecule or bond can comprise a peptidomimetic entity, e.g., a peptidomimetic molecule or bond.

As used herein, the term "altering the spatial position" refers to changing the orientation of or translating the structure of the candidate compound, relative to a pre-defined reference, e.g., relative to the structure of the E6bp molecule. For example, the structure of the candidate compound can be rotated, e.g., 30, 60, 90, 120 or 180° relative to the structure of the E6bp molecule.

As used herein, the term "atom equivalencies" refers to a set of conserved residues between two structures, defined such that they allow direct comparison of the structures being compared. For example, the atom equivalencies can correspond to protein backbone atoms, e.g., N, Cα, C and O atoms.

As used herein, the term "fitting operation" refers to the process by which, a working structure (i.e. a compound) is translated and rotated to obtain an optimum fit with the target E6bp structure. The fitting operation can use a least squares fitting algorithm that computes the optimum translation and rotation to be applied to the moving compound structure, such that the root mean square difference of the fit over the specified pairs of equivalent atoms is an absolute minimum. This number, given in angstroms, can be reported by a computer software.

As used herein, the term "root mean square deviation" refers to the square root of the arithmetic mean of the squares of the deviations from the mean. It is a way to express the deviation or variation from a trend or object. For purposes of this invention, the "root mean square deviation" defines the variation in a set of atom equivalencies of a compound from a set of atom equivalencies of the E6bp molecule, as defined by the structure coordinates of the E6bp molecule described herein.

As used herein, the term "least squares" refers to a method based on the principle that the best estimate of a value is that in which the sum of the squares of the deviations of observed values is a minimum.

The methods of the invention allow rapid and efficient design and evaluation of compounds useful for inhibiting infection by papillomavirus.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

DETAILED DESCRIPTION

BRIEF DESCRIPTION OF THE DRAWINGS

The drawings are first briefly described.

Figure 1 is a depiction of the three-dimensional structure of E6bp. The protein backbone is shown as a blue ribbon. Hydrophobic amino acid side chains are shown in yellow, and polar side-chains in blue. Calcium ions are shown as spheres.

Figure 2 is an amide-amide region of the 2D NOESY illustrating the structure of the E6bp molecule. Intense amide to amide proton contacts, indicative of α helices, are indicated by residue number. Residues 13 to 16, shown above the diagonal, form an N-terminal helix, whereas residues 22 to 27, shown below the diagonal, for a C-terminal helix.

Figure 3 is an illustration of ligand design. In this example, the candidate compound phenylarginine was built onto two exposed amino acids residues (glutamate 16 and leucine 19) on the C-terminal α helix of the E6bp protein. The hydrocarbons of the ligand are indicated by G, whereas those of the protein are indicated by Y. Oxygen atoms are indicated by R, nitrogens by B, and polar hydrogens by W. The compound was designated with the LUDI feature feature of the molecular modeling program INSIGHTII, and the fit to the protein was optimized using the DOCK module. This figure illustrates the method for design of novel inhibitors to papillomavirus, described herein.

Figure 4 is a depiction of various sequences illustrating that the E6 binding domains is a short α-helical peptide and that the E6 binding region of E6BP is found in other E6-binding proteins. Further illustrated in this Figure are the results from an analysis of the structure of E6bp, based on site-directed mutagenesis.

E6 Signal Transduction Pathway

E6-induced tumorigenesis occurs via two pathways. In the p53-dependent pathway, the E6 protein associates with a human cellular factor, E6AP. The E6-E6AP complex directs p53 for rapid degradation via the ubiquitin-mediated proteolytic pathway (Lee J.M. and Bernstein A. (1995) *Cancer Metas. Rev.* 14:149; Scheffner M. et al., (1993) *Cell* 75:495; and Huibregste J. et al., (1993) *Mol. Cell Biol.* 13:775). Loss of p53 protein correlates with the loss in its tumor suppressor functions. An 18 amino acid residue peptide fragment, E6ap, is the minimal region of E6AP that binds E6 (Huibregste J. et al., (1993) *Mol. Cell Biol.* 13:4918).

E6-induced tumorigenesis is also p53-independent (Storey A. et al., (1995) *Oncogene* 11:653), and a different target protein, ERC55, has been demonstrated to bind E6 (Chen J.J. et al., (1995) *Science* 269:529). ERC55 and p53 compete for binding to E6, consistent with alternate roles in tumorigenesis. The mechanism of ERC55 function appears to involve alteration of keratinocyte differentiation (Sherman L. and Schlegel R.

(1996) *J. Virol.* 70:3269; and Reiss M. et al., (1989) *Cancer Commun.* 1:75; Chen J.J. et al., (1995) *Science* 269:529; and Howley P.M. (1996) *Field Virol.* Chapter 65, 2nd edition). A 25 amino acid segment of ERC55, called E6bp, has been found to be necessary and sufficient for binding to E6.

The sequence of E6ap is homologous to E6bp, as shown in the following Table.

TABLE I

E6ap	IPESSELTLQELLGEERR (SEQ I.D. NO:1)
	**|*|*||* |
E6bp	ALEEHDKNGDGFVSLEEFLGDYRWD (SEQ I.D. NO:2)

The * symbol indicates homologous amino acids, whereas the | symbol denotes identical amino acids.

Structural Analysis of E6bp

The inventors have solved the three-dimensional structure of an E6bp molecule, using one-and two-dimensional NMR Spectroscopy. Importantly, this has provided, for the first time, information about the three-dimensional structure of the E6bp molecule.

This information is of significant utility in fields such as drug discovery. An understanding of the structure of the E6bp molecule, a 25 amino acid region in the ERC55 protein sufficient for binding to the HPV E6 transforming protein, allows the design of drugs which interact with the HPV E6 transforming protein. As a result, this information is useful for designing inhibitors of the E6-ERC55 interaction and therefore, drugs for fighting papillomavirus infection.

Candidate Compounds

Candidate compounds can be agents which can be altered so as to have a three-dimensional structure that is substantially similar to the three-dimensional structure of E6bp, provided in Figure 1, such that the agent can bind to the HPV E6 transforming protein or portion thereof. Preferably, the altered candidate compound can bind to E6 with an affinity which is at least 10, 50, 100, 150, 200 or 500% as strong as the affinity with which E6bp binds to E6. Candidate compounds can also be agents which can be altered so as to have a three-dimensional structure that is suitable for associating with the E6bp molecule.

For example, the candidate compound can be a peptide or a peptidomimetic. Examples of peptidomimetics include peptidic compounds in which the peptide backbone is substituted with one or more benzodiazepine molecules (see *e.g.*, James,

G.L. *et al.*, 1993, *Science* 260:1937-1942, the contents of which are incorporated herein by reference), peptides in which all L-amino acids are substituted with the corresponding D-amino acids, and "retro-inverso" peptides (see U.S. Patent No. 4,522,752 by Sisto, the contents of which are incorporated herein by reference).

The term mimetic, and in particular, peptidomimetic, includes isosteres. The term "isostere" as used herein, includes a chemical structure that can be substituted for a second chemical structure because the steric conformation of the first structure fits a binding site specific for the second structure. The term specifically includes peptide back-bone modifications (*e.g.*, amide bond mimetics). Such modifications include modifications of the amide nitrogen, the α-carbon, amide carbonyl, complete replacement of the amide bond, extensions, deletions or backbone cross links. Several peptide backbone modifications are known, including ψ[CH_2S], ψ[CH_2NH], ψ [$CSNH_2$], ψ[NHCO], ψ[$COCH_2$], and ψ[(E) or (Z) CH=CH]. In the nomenclature used above, ψ indicates the absence of an amide bond. The structure that replaces the amide group is specified within the brackets.

Other possible modifications include an N-alkyl (or aryl) substitution (ψ [CONR]), backbone cross linking to construct lactams and other cyclic structures, substitution of all D-amino acids for all L-amino acids within the candidate compound ("inverso" compounds) or retro-inverso amino acid incorporation (ψ[NHCO]). By "inverso" is meant replacing L-amino acids of a sequence with D-amino acids, and by "retro-inverso" or "enantio-retro" is meant reversing the sequence of the amino acids ("retro") and replacing the L-amino acids with D-amino acids. For example, if the parent peptide is Thr-Ala-Tyr, the retro modified form is Tyr-Ala-Thr, the inverso form is thr-ala-tyr, and the retro-inverso form is tyr-ala-thr (lower case letters refer to D-amino acids). Compared to the parent peptide, a retro-inverso peptide has a reversed backbone while retaining substantially the original spatial conformation of the side chains, resulting in a retro-inverso isomer with a topology that closely resembles the parent peptide. (See Goodman *et al.* "*Perspectives in Peptide Chemistry*" pp. 283-294, 1981, and U.S. Patent No. 4,522,752, the contents of which are incorporated herein by reference).

The candidate compound can also be a non peptide organic compound prepared as described in WO 9504277, the contents of which are incorporated herein by reference, as well as a steroid, a carbohydrate, a lipid and the like.

The candidate compound can be selected from a database of three-dimensional structures of known compounds. The three-dimensional structures in the database can be either experimentally determined, e.g., crystal structures from the Cambridge

structural database (see Allen et al., J. Chem. Inf. Comput. Sci. 31: 187-204, 1991, the contents of which are incorporated herein by reference) or computationally generated, e.g., using rule-based programs such as CONCORD (see Pearlman, R. S., Chem. Des. Auto. News, 2:1-7, 1987, the contents of which are incorporated herein by reference).

The candidate compounds can also be designed de novo; e.g., by piecing together or assembling molecular fragments to create compounds which: (a) have a three-dimensional structure that is substantially similar to the three-dimensional structure of E6bp provided in figure 1, such that the created compound can bind to the HPV E6 transforming protein or portion thereof or (b) have a three-dimensional structure that is suitable for associating with the E6bp molecule. For example, the GROW algorithm (Moon, J.B., et al., *Proteins: Struct. Funct. Genet* 11:314-328,1991, the contents of which are incorporated herein by reference), or the LUDI program (Böhm, H.-J. *J Comput Aided Mol Design* 6:61-78,1992, the contents of which are incorporated herein by reference) can be used.

Other useful programs to aid one of skill in the art in assembling molecular components to create compounds include: (1) CAVEAT, described in Bartlett et al., In "Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc., 78, 182-196,1989, the contents of which are incorporated herein by reference; (2) 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, CA). The use of these systems is described in Martin et al., J. Med. Chem., 35, 2145-2154,1992, the contents of which are incorporated herein by reference; and (3) HOOK (available from Molecular Simulations, Burlington, MA).

Other molecular modeling techniques may also be employed in accordance with this invention. See e.g., Cohen et al., J. Med. Chem., 33, 883-894,1990, Naiva et al., Current Opinions in Structural Biology, 2, 202-210,1992, the contents of which are incorporated herein by reference.

Once a compound has been designed by the above methods, its similarity to the three-dimensional structure of the E6bp peptide may be evaluated.

Machine Readable Storage Medium

In order to use the NMR derived structure coordinates for the E6bp peptide, it is preferable to convert them into a three-dimensional representation. This can be achieved through the use of commercially available software which is capable of generating three-dimensional graphical representations of molecules or portions thereof from as set of structure coordinates.

Evaluation and Design of Candidate Compounds

The invention allows the use of molecular design techniques to design and evaluate candidate compounds, including inhibitory compounds, e.g., candidate compounds having a three-dimensional structure that is: (a) substantially similar to the three-dimensional structure of E6bp provided in figure 1, such that the candidate compounds can bind to the HPV E6 transforming protein or portion thereof or (b) suitable for associating with the E6bp molecule.

A potential compound which can bind to the HPV E6 transforming protein or portion thereof can be evaluated by means of a series of steps in which compounds are screened and selected for their similarity to the three-dimensional structure of the E6bp molecule.

One skilled in the art can use one of several methods to screen compounds for their similarity to the three-dimensional structure of the E6bp molecule. This process may begin by visual inspection of, for example, the three-dimensional structure of the candidate compound in comparison to the three-dimensional structure of the E6bp molecule on a computer screen, wherein the three-dimensional structure of the E6bp molecule is generated from the machine-readable storage medium.

Various computational analyses can also be used to determine whether a compound is sufficiently similar to the three-dimensional structure of the E6bp molecule. Such analyses can be carried out in current software applications, such as the Molecular Similarity application of QUANTA (Molecular Simulations Inc., Waltham, MA) version 3.3, and as described in the accompanying User's Guide, Volume 3 pg. 134-135, the contents of which are incorporated herein by reference.

The Molecular Similarity application permits comparisons between different structures, different conformations of the same structure, and different parts of the same structure. The procedure used in Molecular Similarity to compare structure is divided into four steps: 1) loading the structures to be compared; 2) defining the atom equivalences in these structures; 3) performing a fitting operation; and 4) analyzing the results.

Each structure can be identified by a name. The E6bp structure can be identified as the target (i.e., the fixed structure); the candidate compound structures can be working structures (i.e., moving structures). Since atom equivalency within QUANTA is defined by user input, equivalent atoms can be defined as protein backbone atoms (N, Cα, C and O) for all conserved residues between the two structures being compared. The process can be aided by color-coding the different parts of the molecules.

Rigid fitting operations can be used. When a rigid fitting method is used, the working compound structure is translated and rotated to obtain an optimum fit with the

target E6bp structure. The fitting operation uses a least squares fitting algorithm that computes the optimum translation and rotation to be applied to the moving compound structure, such that the root mean square difference of the fit over the specified pairs of equivalent atom is an absolute minimum. This number, given in angstroms, can be reported by QUANTA.

Preferred candidate structures are those having a set of structure coordinates with a root mean square deviation of conserved residue backbone atoms (e.g., N, Cα, C, O) of less than 1.5 Å when superimposed, using backbone atoms, on the relevant structure coordinates listed in Figure 2 are considered identical. More preferably, the root mean square deviation is less than 1.0 Å, and even more preferably, the root mean square deviation is less than 0.5 Å.

Candidate compounds can also be evaluated for their ability to associate with the E6bp molecule. One skilled in the art may use one of several methods to screen compounds for their ability to associate with the E6bp molecule. This process may begin by visual inspection of, for example, the E6bp molecule on a computer screen based on the NMR derived data shown in Figure 2, and generated from the machine-readable storage medium. Selected compounds may then be positioned in a variety of orientations, or docked, within the E6bp molecule. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics force fields, such as CHARMM and AMBER. Specialized computer programs may also assist in the process of selecting fragments or chemical entities. These include: 1. AUTODOCK (D.S. Goodsell et al., "Automated Docking of Substrates to Proteins by Simulated Annealing" Proteins: Structure, Function, and Genetics, 8, pp. 195-202 (1990)). DOCK is available from University of California, San Francisco, CA. 2. DOCK, (I.D. Kuntz et al., J. Mol. Biol., 161, pp. 269-288 (1982), the contents of which are incorporated herein by reference). DOCK is available from University of California, San Francisco, CA.

Synthesis of Identified Compounds

Once a compound has been evaluated by the above methods, it can be prepared by standard techniques known in the art. Peptides can be synthesized using standard techniques such as those described in Bodansky, M. Principles of Peptide Synthesis, Springer Verlag, Berlin (1993) and Grant, G.A (ed.). Synthetic Peptides: A User's Guide, W.H. Freeman and Company, New York (1992). Automated peptide synthesizers are commercially available (*e.g.*, Advanced ChemTech Model 396; Milligen/ Biosearch 9600). Approaches to designing peptide analogs are also known in the art. For example, see Farmer, P.S. in Drug Design (E.J. Ariens, ed.) Academic

Press, New York, 1980, vol. 10, pp. 119-143; Ball. J.B. and Alewood, P.F. (1990) *J. Mol. Recognition* 3:55; Morgan, B.A. and Gainor, J.A. (1989) *Ann. Rep. Med. Chem.* 24:243; and Freidinger, R.M. (1989) *Trends Pharmacol. Sci.* 10:270, the contents of all of which are incorporated herein by reference.

Alternatively, peptide compounds can be prepared according to standard recombinant DNA techniques using a nucleic acid molecule encoding the peptide. A nucleotide sequence encoding the peptide can be determined using the genetic code and an oligonucleotide molecule having this nucleotide sequence can be synthesized by standard DNA synthesis methods (*e.g.*, using an automated DNA synthesizer). Alternatively, a DNA molecule encoding a peptide compound can be derived from the corresponding natural gene or cDNA (*e.g.*, using the polymerase chain reaction and/or restriction enzyme digestion) according to standard molecular biology techniques.

To facilitate expression of a peptide compound in a host cell by standard recombinant DNA techniques, the isolated nucleic acid encoding the peptide is incorporated into a recombinant expression vector. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" or simply "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, such other forms of expression vectors, such as viral vectors, which serve equivalent functions may also be used to express a peptide compound.

The nucleotide sequence encoding the peptide compound can be operatively linked to one or more regulatory sequences, selected on the basis of the host cells to be used for expression. The term "operably linked" is intended to mean that the sequences encoding the peptide compound are linked to the regulatory sequence(s) in a manner that

allows for expression of the peptide compound. The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990), the content of which are incorporated herein by reference. Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cells, those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences) and those that direct expression in a regulatable manner (*e.g.*, only in the presence of an inducing agent). It will be appreciated by those skilled in the art that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed, the level of expression of peptide compound desired, and the like. The peptide compound expression vectors can be introduced into host cells to thereby produce peptide compounds encoded by nucleic acids.

The recombinant expression vectors can be designed for expression of peptide compounds in prokaryotic or eukaryotic cells. For example, peptide compounds can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector may be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.*, (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Baculovirus vectors available for expression of peptide compounds in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, (1983) *Mol. Cell. Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39).

Examples of mammalian expression vectors include pCDM8 (Seed, B., (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987), *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

In addition to the regulatory control sequences discussed above, the recombinant expression vector may contain additional nucleotide sequences. For example, the recombinant expression vector may encode a selectable marker gene to

identify host cells that have incorporated the vector. Such selectable marker genes are well known in the art. Moreover, to facilitate secretion of the peptide compound from a host cell, in particular mammalian host cells, the recombinant expression vector preferably encodes a signal sequence operatively linked to sequences encoding the amino-terminus of the peptide compound such that upon expression, the peptide compound is synthesized with the signal sequence fused to its amino terminus. This signal sequence directs the peptide compound into the secretory pathway of the cell and is then cleaved, allowing for release of the mature peptide compound (*i.e.*, the peptide compound without the signal sequence) from the host cell. Use of a signal sequence to facilitate secretion of proteins or peptides from mammalian host cells is well known in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells *via* conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, microinjection and viral-mediated transfection. Suitable methods for transforming or transfecting host cells can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene encoding the peptide compound. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker may be introduced into a host cell on the same vector as that encoding the peptide compound or may be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

Assessment of the Biological Activity of Identified Compounds

Once a compound has been identified and synthesized, its biological activity can be assessed. For example, an assay can be used to assess the ability of a compound to inhibit binding between ERC55 (i.e., E6bp) and the HPV E6 transforming protein. A variety of assays are available and readily apparent to the skilled artisan. For example,

in one such biological evaluation assay the identified compound can be contacted with an isolated and purified ERC55 protein. The mixture of the candidate compound and ERC55 can then be added to a composition containing the E6 protein but which does not contain ERC55. Detection and quantification of labelled E6/ERC55 complexes provides a means for determining the candidate compound's efficacy at inhibiting complex formation between the papillomavirus E6 protein and the ERC55 protein. A control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified ERC55 is added to a composition containing the E6 protein, and the formation of an E6/ERC55 complex is quantitated in the absence of the test compound.

Complex formation between ERC55 and E6 may be detected by a variety of other methods as well. For example, glutathione-S-transferase/E6 (GST/E6) fusion proteins can be adsorbed onto glutathione sepharose beads which can then be combined with an ^{35}S-labeled ERC55 protein and incubated under conditions conducive to complex formation, [e.g., at 4°C in a buffer of 25 mM Tris-HCl (pH 7.2), 50 mM NaCl and 0.2% NP-40]. Following incubation, the beads can be washed to remove any unbound ERC55, and the sepharose bead-bound radiolabel can be determined directly (e.g. beads placed in scintilant), or in the supernatant after the E6/ERC55 complexes are dissociated (e.g. by treatment with DTT). The supernatant containing the complexes can be separated by SDS-PAGE gel before detection.

Additionally, ERC55 or E6bp can be used to generate a two-hybrid assay, as described in U.S. Patent No: 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J Biol Chem* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; and Iwabuchi et al. (1993) *Oncogene* 8:1693-1696), for subsequently assessing the ability of compounds to disrupt binding of ERC55 or E6bp to E6. The interaction trap assay relies on reconstituting *in vivo* a fuctional transcriptional activator protein from two separate fusion proteins, one of which comprises the DNA-binding domain of a transcriptional activator fused to an E6 protein. The second fusion protein comprises a transcriptional activation domain (e.g. able to initiate RNA polymerase transcription) fused to ERC55 or E6bp. When the E6 and ERC55 or E6bp proteins interact, the two domains of the transcriptional activator protein are brought into sufficient proximity as to cause transcription of a reporter gene. For example, *Saccharomyces cerevisiae* YPB2 cells can be transformed simultaneously with a plasmid encoding a GAL4db-E6 fusion and with a plasmid encoding the GAL4ad domain fused to ERC55 or E6bp. Moreover, the strain is transformed such that the GAL4-responsive promoter drives expression of a phenotypic marker. For example, the ability to grow in the absence of histidine depends

on the expression of the HIS3 gene. When the HIS3 gene is placed under the control of a GAL4-responsive promoter, relief of this auxotrophic phenotype indicates that a functional GAL4 activator has been reconstituted through the interaction of E6 and ERC55 or E6bp. Thus, a compound able to inhibit an ERC55 or E6bp interaction with E6 will result in yeast cells unable to grow in the absence of histidine. Alternatively, the phenotypic marker (e.g. instead of the HIS3 gene) can be one which provides a negative selection when expressed such that compounds which disrupt E6/ERC55 or E6/E6bp interactions confer positive growth selection to the cells.

EXAMPLES

Example 1: Sample Preparation

The E6bp peptide was synthesized using standard protein synthesis methods and purified by HPLC. The recombinant E6 protein is expressed in *E. coli*, solubilized from inclusion bodies, and refolded using a rapid dilution method. The E6 protein has been expressed and solubilized from HPV-16 to a concentration of 4mg/mL (0.2 mM). For the peptide-E6 complex, (protonated) peptide is added to 0.2 mM deuterated (^{2}H)E6 following the procedure of Shibata et al., (1995) *Arch. Biochem. Biophys.* 319:204, the contents of which are incorporated herein by reference. This allows observation of only the ^{1}H NMR resonances of the E6-bound peptides.

Example 2: NMR Spectroscopy

Peptide samples were prepared at about 2 mM concentration. One- and two-dimensional (1D & 2D) NMR data were collected on a 500 MHz Bruker AMX500 spectrometer. E6bp is titrated with calcium and NMR spectra recorded to determine the effect on the peptide conformation. Low concentration peptide samples (0.2 mM) are titrated with E6 to determine the residues that are affected most by E6. Samples are prepared in both H_2O and D_2O solution.

Since the equilibrium binding constant of E6bp-E6 complex is in the μM range, transferred NOE (TRNOE) experiments, as described in Clore M. and Gronenborn A.M., (1982) *J. Magn. Reson.* 48:402; *ibid.* (1983) 53:423; and Sykes B.D., (1994) *Cur. Opin. Biotech.* 4:392, the contents of which are incorporated herein by reference, are suitable to determine the conformation of E6-bound E6bp. TRNOE experiments are performed with an approximately 10-fold excess of ligand relative to the protein (2 mM peptide and 0.2 mM E6 protein). In TRNOE experiments, the spectra exhibit mostly the lineshape and intensity of the unbound peptide, but the NOESY cross peaks represent the bound conformation of the peptide. Spectra are optimized in terms of salt concentration (up to 0.2 M) and temperatures (5°C to 35°C) to identify the conditions

that yield superior quality NMR data with narrow line shapes, following the methods described in Baleja J.D., (1996) *Techniques in Protein Chemistry* VII:131, the contents of which are incorporated herein by reference.

Total correlation spectroscopy (TOCSY), and nuclear Overhauser effect spectroscopy (NOESY) experiments are used for resonance assignment and to obtain conformation data for structure calculations, as described in Bax A. and Davis D.G., (1985) *J. Magn. Reson.* 65:355; and Wüthrich K., (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, the contents of which are incorporated herein by reference. NMR data are processed using the Bruker NMR processing program or FELIX (Biosym, Inc.)

Example 3: NMR Resonance Assignment Strategy

NMR resonance strategy is performed as described in Wüthrich et al. In this method, TOCSY spectra are analyzed to provide information regarding spin-spin coupled NMR-active nuclei. Different types of amino acids side chains produce distinct TOCSY cross peak patterns. The cross-peak patterns for each amino acid are linked together using the NOESY experiment since NOE sequential connectivities can be identified between residues neighboring in sequence. NOESY spectra show cross peaks between protons that are within 5 Å of each other and therefore also show contacts between residues that are sequentially distant but spatially close. Interproton distances are determined from NOESY cross peak intensities. Φ and χ-1 torsion angles are obtained by measuring the coupling constants from 1D slices of resolution-enhanced 2D data, as described in Wüthrich K., (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York; and Szyperski T. et al., (1992) *J. Magn. Reson.* 99:552, the contents of which are incorporated herein by reference.

^{1}H NMR resonance assignments were made (see Table II) and a low resolution structure of the calcium-bound E6bp was calculated using approximately 200 interproton distances and 15 Φ torsion angles derived from NOESY data and distance geometry and simulated annealing protocols of the INSIGHTII molecular modeling program (see Figure 1). Consistent with peptide analogs of other EF-hand proteins, the E6bp peptide dimerizes. From the dispersion present in the NMR spectra, it is estimated that 400 additional interproton distances are obtainable. Structure calculation is repeated with the more extensive data set to determine the high resolution structure of E6bp. The high resolution structure of E6bp both free and bound to E6 protein is also solved. Knowing the conformation for this E6bp in the absence of E6 protein is important for understanding the conformational changes brought about by binding to E6 and the binding surface of the peptide involved in the interaction.

TABLE II

NMR Resonance Assignments for E6 binding peptide

Residue	HN	HA	HB	HG	Others
Glu 1	-	4.02	2.08,2.08		2.37,2.37
Phe 2	8.10	4.76	3.08,3.03		D, 7.21; E, 7.15; Z, 7.28*
Val 3	8.11	4.16	1.95	0.89,0.89	
Ile 4	8.11	3.98	1.79	1.46,1.12	G2, 0.86 D1, 0.80
Gln 5	8.35	4.15	2.10,2.06	2.37,2.37	E, 7.51,6.76
Glu 6	8.00	4.22	2.08,2.02*	2.29,2.29	
Ala 7	7.98	4.35	1.46		
Leu 8	8.12	4.17	1.69,1.62	1.69	D, 0.81,0.79
Glu 9	8.22	3.98	2.08,2.07	2.36,2.34	
Glu 10	7.98	4.11	1.87,1.84	2.12,2.12	
His 11	8.03	4.63	3.33,3.12		D, 7.38; E, 8.30
Asp 12	7.73	4.76	3.00,2.37		
Lys 13	7.95	4.15	1.97,1.92	1.64,1.56	D, 1.70,1.70; E, 3.09,3.09
Asn 14	8.17	4.83	3.31,2.91		D, 7.97,6.75
Gly 15	7.66	3.92,3.84			
Asp 16	8.06	4.56	3.17,2.53		
Gly 17	9.93	4.10,3.48			
Phe 18	8.05	4.90	2.95,2.88		D, 7.12; E, 7.26; Z, 7.38*
Val 19	8.78	4.70	1.96	0.93,0.83	
Ser 20	9.11	4.72	4.49,4.05		
Leu 21	8.77	3.99	1.67,1.52	1.50*	D, 0.73,0.71
Glu 22	8.85	3.92	2.07,1.98	2.37,2.29	
Glu 23	7.79	4.05	2.70,2.55	2.62,2.33	
Phe 24	8.30	4.08	3.03,2.95		D, 6.72; E, 6.88; Z,6.89*
Leu 25	8.56	4.03	1.78,1.50	1.50	D, 0.72,0.69
Gly 26	7.78	3.92,3.92*			
Asp 27	7.71	4.74	2.70,2.55		
Tyr 28	7.73	4.32	2.78,2.70		D, 6.80; E, 6.62
Arg 29	7.67	4.25	1.64,1.54	1.38	D, 3.04,3.04; E, 7.31, H, -
Trp 30	7.88	4.60	3.32,3.18		D, 7.25; E1, 10.01; E3, 7.58 Z2, 7.35; Z3, 7.08; H, 7.08
Asp 31	8.09	4.53	2.55,2.44		
NH2	7.56*				6.86*

* Tentative assignment. "-" not observable. 1 mM peptide was in a buffer containing 100 mM NaCl, 50 mM calcium chloride, pH 6.0, 10% TFE, 35 deg. C.

Structure Calculation and Analysis

The interproton distances and torsion angle information (see Table III) are introduced into the distance geometry program, DGII of Insight (Biosym, Inc.). The initial structure calculated using distance geometry and simulated annealing protocols is refined by determining the presence or lack of NOE cross peaks for spatially proximal protons in the calculated structure.

Structures are compared to one another to highlight the conformational changes. For example, the E6bp peptide is expected to have major changes in surface properties on binding calcium, such as the formation of a deep hydrophobic cavity essential for target protein recognition (see Ikura M., (1996) *Trends Biochem. Sci.* 21:14, the contents of which are incorporated herein by reference). The structures of E6ap and E6bp are compared to one another to test the hypothesis that they are similar. The residues involved in interaction with E6 are identified by chemical shift changes that occur upon titration, as described in Baleja J.D. et al., (1992) *Nature* 356:450; and Baleja J.D. et al., (1992) *Biochemistry* 33:3071, the contents of which are incorporated herein by reference. Furthermore, differences in conformation observed on binding E6 are determined.

The high resolution structures and the identification of key residues for binding E6 are important for the design of small molecule inhibitors that block the oncogenic properties of E6 (see Fesik S.W., (1993) *J. Biomol. NMR* 3:261; and Kuntz I.D. et al., (1994) *Acc. Chem. Res.* 27:117, the contents of which are incorporated herein by reference). The first focus is on readily synthesized peptide derivatives for *in vitro* testing. Modifications of a lead compound to optimize interactions are proposed using the BUILDER and LUDI features of the INSIGHTII molecular modeling program (and chemical intuition). Binding to the protein surface uses the DOCK module. An illustration of the approach is given in Figure 3.

TABLE III

STRUCTURE DATA FOR E6 BINDING PEPTIDE

BIOSYM restraint 1

remote prochiral center

VAL_19:HG1*	VAL_19:HG2*	VAL_19:CG1	VAL_19:CG2	VAL_19:CB
VAL_19B:HG1*	VAL_19B:HG2*	VAL_19B:CG1	VAL_19B:CG2	VAL_19B:CB
LEU_21:HD1*	LEU_21:HD2*	LEU_21:CD1	LEU_21:CD2	LEU_21:CG
LEU_21B:HD1*	LEU_21B:HD2*	LEU_21B:CD1	LEU_21B:CD2	LEU_21B:CG

NOE distance

Distance between atoms (interproton distances) within the E6 binding peptide

ASP-_12:OD1	CAL_40:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
ASP-_12B:OD1	CAL_40B:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
ASN_14:OD1	CAL_40:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
ASN_14B:OD1	CAL_40B:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
ASP-_16:OD1	CAL_40:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
ASP-_16B:OD1	CAL_40B:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
GLU-_23:OE1	CAL_40:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
GLU-_23B:OE1	CAL_40B:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
GLU-_23:OE2	CAL_40:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00

GLU-_23B:OE2	CAL_40B:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
PHE_18:O	CAL_40:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
PHE_18B:O	CAL_40B:CA6	2.000	2.720	2.800	1.00	1.00	1000.000	0.00
ILE_4:HN	GLN_5:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ILE_4B:HN	GLN_5B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ILE_4:HB	GLN_5:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ILE_4B:HB	GLN_5B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLN_5:HN	GLU-_6:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLN_5B:HN	GLU-_6B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLN_5:HB*	GLU-_6:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
GLN_5B:HB*	GLU-_6B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
GLN_5:HG*	GLU-_6:HN	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
GLN_5B:HG*	GLU-_6B:HN	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
GLU-_6:HN	ALA_7:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_6B:HN	ALA_7B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_6:HG*	ALA_7:HN	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
GLU-_6B:HG*	ALA_7B:HN	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HN	LEU_8:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HN	LEU_8B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HB*	LEU_8:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HB*	LEU_8B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LEU_8:HN	GLU-_9:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LEU_8B:HN	GLU-_9B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_9:HN	GLU-_10:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_9B:HN	GLU-_10B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
HIS_11:HN	ASP-_12:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
HIS_11B:HN	ASP-_12B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ASP-_12:HN	LYS+_13:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ASP-_12B:HN	LYS+_13B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
LYS+_13:HN	ASN_14:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LYS+_13B:HN	ASN_14B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
ASN_14:HN	GLY_15:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
ASN_14B:HN	GLY_15B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
ASN_14:HN	GLY_15:HA1	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
ASN_14B:HN	GLY_15B:HA1	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
ASN_14:HN	GLY_15:HA2	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
ASN_14B:HN	GLY_15B:HA2	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
GLY_15:HN	ASP-_16:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLY_15B:HN	ASP-_16B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLY_15:HA1	ASP-_16:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
GLY_15:HA2	ASP-_16:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
GLY_15B:HA1	ASP-_16B:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
GLY_15B:HA2	ASP-_16B:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
ASP-_16:HN	GLY_17:HN	2.000	2.900	2.800	1.00	1.00	1000.000	0.00
ASP-_16B:HN	GLY_17B:HN	2.000	2.900	2.800	1.00	1.00	1000.000	0.00
ASP-_16:HA	GLY_17:HN	3.000	5.000	2.800	1.00	1.00	1000.000	0.00
ASP-_16B:HA	GLY_17B:HN	3.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLY_17:HA1	PHE_18:HN	2.000	2.500	2.800	1.00	1.00	1000.000	0.00
GLY_17:HA2	PHE_18:HN	2.000	2.500	2.800	1.00	1.00	1000.000	0.00
GLY_17B:HA1	PHE_18B:HN	2.000	2.500	2.800	1.00	1.00	1000.000	0.00
GLY_17B:HA2	PHE_18B:HN	2.000	2.500	2.800	1.00	1.00	1000.000	0.00
PHE_18:HA	VAL_19:HN	2.000	2.500	2.800	1.00	1.00	1000.000	0.00
PHE_18B:HA	VAL_19B:HN	2.000	2.500	2.800	1.00	1.00	1000.000	0.00
LEU_21:HN	GLU-_22:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LEU_21B:HN	GLU-_22B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LEU_21:HB*	GLU-_22:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
LEU_21B:HB*	GLU-_22B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00

GLU-_22:HN	GLU-_23:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22B:HN	GLU-_23B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22:HB*	GLU-_23:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
GLU-_22B:HB*	GLU-_23B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
GLU-_23:HN	PHE_24:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_23B:HN	PHE_24B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
GLU-_23:HB*	PHE_24:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
GLU-_23B:HB*	PHE_24B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
PHE_24:HN	LEU_25:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
PHE_24B:HN	LEU_25B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
PHE_24:HB*	LEU_25:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
PHE_24B:HB*	LEU_25B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
LEU_25:HN	GLY_26:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LEU_25B:HN	GLY_26B:HN	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
LEU_25:HB*	GLY_26:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
LEU_25B:HB*	GLY_26B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
GLY_26:HN	ASP-_27:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
GLY_26B:HN	ASP-_27B:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
ASP-_27:HN	TYR_28:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
ASP-_27B:HN	TYR_28B:HN	2.000	3.500	2.800	1.00	1.00	1000.000	0.00
TYR_28:HN	ARG+_29:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
TYR_28B:HN	ARG+_29B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
TYR_28:HB*	ARG+_29:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
TYR_28B:HB*	ARG+_29B:HN	2.000	4.500	2.800	1.00	1.00	1000.000	0.00
ARG+_29:HN	TRP_30:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ARG+_29B:HN	TRP_30B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ARG+_29:HB*	TRP_30:HN	3.000	6.000	2.800	1.00	1.00	1000.000	0.00
ARG+_29B:HB*	TRP_30B:HN	3.000	6.000	2.800	1.00	1.00	1000.000	0.00
ARG+_29:HG*	TRP_30:HN	3.000	6.000	2.800	1.00	1.00	1000.000	0.00
ARG+_29B:HG*	TRP_30B:HN	3.000	6.000	2.800	1.00	1.00	1000.000	0.00
ARG+_29:HA	TRP_30:HN	2.000	2.600	2.800	1.00	1.00	1000.000	0.00
ARG+_29B:HA	TRP_30B:HN	2.000	2.600	2.800	1.00	1.00	1000.000	0.00
TRP_30:HN	AP-C_31:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
TRP_30B:HN	AP-C_31B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ILE_4:HA	ALA_7:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ILE_4B:HA	ALA_7B:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLN_5:HN	ALA_7:HB*	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLN_5B:HN	ALA_7B:HB*	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HA	GLU-_10:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HA	GLU-_10B:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HA	HIS_11:HD2	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HA	HIS_11B:HD2	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HB*	GLU-_9:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HB*	GLU-_9B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HB*	GLU-_10:HG*	2.000	8.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HB*	GLU-_10B:HG*	2.000	8.000	2.800	1.00	1.00	1000.000	0.00
GLU-_9:HA	HIS_11:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLU-_9B:HA	HIS_11B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
HIS_11:HB*	VAL_19:HG1*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
HIS_11B:HB*	VAL_19B:HG1*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
HIS_1HB*	VAL_19:HG2*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
HIS_11B:HB*	VAL_19B:HG2*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ASN_14:HD21	ASP-_16:HB*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ASN_14B:HD21	ASP-_16B:HB*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
GLY_15:HA1	GLY_17:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLY_15:HA2	GLY_17:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLY_15B:HA1	GLY_17B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00

GLY_15B:HA2	GLY_17B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
SER_20:HN	GLU-_23:HN	2.000	5.900	2.800	1.00	1.00	1000.000	0.00
SER_20B:HN	GLU-_23B:HN	2.000	5.900	2.800	1.00	1.00	1000.000	0.00
SER_20:HN	GLU-_23:HB*	2.000	5.900	2.800	1.00	1.00	1000.000	0.00
SER_20B:HN	GLU-_23B:HB*	2.000	5.900	2.800	1.00	1.00	1000.000	0.00
SER_20:HN	GLU-_23:HG*	2.000	6.900	2.800	1.00	1.00	1000.000	0.00
SER_20B:HN	GLU-_23B:HG*	2.000	6.900	2.800	1.00	1.00	1000.000	0.00
LEU_21:HA	PHE_24:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
LEU_21B:HA	PHE_24B:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
LEU_21:HD1*	PHE_24:HB*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
LEU_21B:HD1*	PHE_24B:HB*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
LEU_21:HD2*	PHE_24:HB*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
LEU_21B:HD2*	PHE_24B:HB*	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22:HA	PHE_24:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22B:HA	PHE_24B:HN	2.000	5.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22:HA	LEU_25:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22B:HA	LEU_25B:HN	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22:HA	LEU_25:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
GLU-_22B:HA	LEU_25B:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
PHE_24:HA	ASP-_27:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
PHE_24B:HA	ASP-_27B:HB*	2.000	4.000	2.800	1.00	1.00	1000.000	0.00
PHE_24:HB*	GLY_26:HN	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
PHE_24B:HB*	GLY_26B:HN	2.000	6.000	2.800	1.00	1.00	1000.000	0.00
LEU_25:HA	TYR_28:HB*	2.000	4.300	2.800	1.00	1.00	1000.000	0.00
LEU_25B:HA	TYR_28B:HB*	2.000	4.300	2.800	1.00	1.00	1000.000	0.00
ALA_7:HA	PHE_24B:CG	3.500	99.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HA	PHE_24:CG	3.500	99.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HA	PHE_24B:CZ	3.500	99.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HA	PHE_24:CZ	3.500	99.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HB*	PHE_24B:CG	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HB*	PHE_24:CG	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ALA_7:HB*	PHE_24B:CZ	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ALA_7B:HB*	PHE_24:CZ	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ILE_4:HD1*	TYR_28B:CG	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ILE_4B:HD1*	TYR_28:CG	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ILE_4:HD1*	TYR_28B:CZ	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
ILE_4B:HD1*	TYR_28:CZ	2.000	7.000	2.800	1.00	1.00	1000.000	0.00
LEU_8:CG	PHE_24B:CG	2.000	9.000	2.800	1.00	1.00	1000.000	0.00
LEU_8B:CG	PHE_24:CG	2.000	9.000	2.800	1.00	1.00	1000.000	0.00
PHE_18:CG	SER_20B:HB*	2.000	9.000	2.800	1.00	1.00	1000.000	0.00
PHE_18B:CG	SER_20:HB*	2.000	9.000	2.800	1.00	1.00	1000.000	0.00
PHE_18:HN	LEU_21B:HB*	3.500	99.000	2.800	1.00	1.00	1000.000	0.00
PHE_18B:HN	LEU_21:HB*	3.500	99.000	2.800	1.00	1.00	1000.000	0.00
PHE_18:HA	SER_20B:HA	2.000	3.000	2.800	1.00	1.00	1000.000	0.00
PHE_18B:HA	SER_20:HA	2.000	3.000	2.800	1.00	1.00	1000.000	0.00

3J dihedral

torsion angle information (theta and chi-1)

LEU_21:HN	LEU_21:N	LEU_21:CA	LEU_21:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
GLU-_22:HN	GLU-_22:N	GLU-_22:CA	GLU-_22:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
PHE_24:HN	PHE_24:N	PHE_24:CA	PHE_24:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LEU_25:HN	LEU_25:N	LEU_25:CA	LEU_25:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LEU_25:C	GLY_26:N	GLY_26:CA	GLY_26:C	5.5	1.0	60.0	60.0	1000.0	-90.0
ASP-_27:HN	ASP-_27:N	ASP-_27:CA	ASP-_27:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
TYR_28:HN	TYR_28:N	TYR_28:CA	TYR_28:HA	5.5	1.0	60.0	60.0	1000.0	-155.6

ARG+_29:HN	ARG+_29:N	ARG+_29:CA	ARG+_29:HA	5.5	1.0	60.0	60.0	1000.0	-155.6
HIS_11:HN	HIS_11:N	HIS_11:CA	HIS_11:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
ASP-_12:HN	ASP-_12:N	ASP-_12:CA	ASP-_12:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LYS+_13:HN	LYS+_13:N	LYS+_13:CA	LYS+_13:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
ASN_14:HN	ASN_14:N	ASN_14:CA	ASN_14:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
ASN_14:C	GLY_15:N	GLY_15:CA	GLY_15:C	5.5	1.0	60.0	60.0	1000.0	-90.0
ASP-_16:HN	ASP-_16:N	ASP-_16:CA	ASP-_16:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LEU_21B:HN	LEU_21B:N	LEU_21B:CA	LEU_21B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
GLU-_22B:HN	GLU-_22B:N	GLU-_22B:CA	GLU-_22B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
PHE_24B:HN	PHE_24B:N	PHE_24B:CA	PHE_24B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LEU_25B:HN	LEU_25B:N	LEU_25B:CA	LEU_25B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LEU_25B:C	GLY_26B:N	GLY_26B:CA	GLY_26B:C	5.5	1.0	60.0	60.0	1000.0	-90.0
ASP-_27B:HN	ASP-_27B:N	ASP-_27B:CA	ASP-_27B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
TYR_28B:HN	TYR_28B:N	TYR_28B:CA	TYR_28B:HA	5.5	1.0	60.0	60.0	1000.0	-155.6
ARG+_29B:HN	ARG+_29B:N	ARG+_29B:CA	ARG+_29B:HA	5.5	1.0	60.0	60.0	1000.0	-155.6
HIS_11B:HN	HIS_11B:N	HIS_11B:CA	HIS_11B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
ASP-_12B:HN	ASP-_12B:N	ASP-_12B:CA	ASP-_12B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
LYS+_13B:HN	LYS+_13B:N	LYS+_13B:CA	LYS+_13B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
ASN_14B:HN	ASN_14B:N	ASN_14B:CA	ASN_14B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6
ASN_14B:C	GLY_15B:N	GLY_15B:CA	GLY_15B:C	5.5	1.0	60.0	60.0	1000.0	-90.0
ASP-_16B:HN	ASP-_16B:N	ASP-_16B:CA	ASP-_16B:HA	5.5	1.0	60.0	60.0	1000.0	-145.6

Structure Analysis of ERC55

Sequence analysis of the E6-interacting protein, ERC55, revealed homology to proteins containing the calcium-binding helix-loop-helix motif called an EF hands. Six EF hands are predicted to occur in its C-terminal domain. Only one of these EF hands, a 25 amino acid segment, E6bp, binds E6 selectively (Chen J.J. et al., (1995) *Science* 269:529), and with about the same affinity as the full-length ERC55 protein. Structural determination of other EF-hand domains using ^{1}H NMR spectroscopy (Ikura M. (1996) *Trends Biochem. Sci.* 21:14) suggests the feasibility for NMR study of this peptide. A small amount of E6bp peptide was synthesized, purified and a 0.7 mM sample was prepared. As predicted, the peptide bound calcium and showed excellent NMR spectral dispersion as shown in Figure 2. The calcium free form of E6bp was studied.

Structure Analysis of E6AP

An 18 amino acid residue peptide fragment, E6ap, is the minimal region of E6AP that binds E6. The sequence of E6ap is homologous to E6bp (see Table I), and their solution properties are likely to be similar. Sequence prediction (Wishart et al., (1994) *Comp. Biol. Sci.* 10:121) has indicated α helical structure for the homologous region of E6ap-the same region which has been determined to be α helical in the primary solution structure of E6bp (Figure 1). E6ap does not have the calcium ligands required for an EF hand, and therefore is unlikely to bind calcium. The small sizes of the E6bp and E6ap peptides make them attractive for detailed structural studies by solution NMR methods in a timely manner.

Example 4: Analysis of The E6bp Structure By Site-Directed Mutagenesis

To further map the domain that is important for E6 binding, additional deletion mutants from E6bp were constructed as GST fusions and tested for binding with E6. Deletion of the first alpha helix (20mer) did not affect binding, nor did deletion of two additional amino acid residues from the C-terminal end of E6bp to the distal region of the second alpha helix affect binding. A 13 amino acid peptide containing the second alpha helix as a GST fusion retained the ability to bind to E6, although at reduced efficiency in comparison to intact E6bp. Notably, the major portion of the loop region from the EP-hand motif is deleted in the 13 amino acid peptide. The ability of the second alpha helix to bind E6 demonstrates that the interaction of E6BP with E6 is independent of calcium binding, as the first alpha helix and loop region from the EP-hand motif are both required for calcium binding.

Alanine replacement mutations in E6bp were also constructed and used to define the amino acid(s) important for E6 interaction in the alpha helix. Some mutations were also made in the surrounding regions. As expected, mutants V19A, S20A, E22A, R29A, and W30A, which have mutations in the area beyond the alpha helix, bound E6 at wild-type level (see Figure 4). Mutant F18A showed some reduced binding. Notably, phenyalanine is a hydrophobic residue, which may enhance the interaction between E6 and E6bp. This may explain why a leucine to alanine change at amino acid 21 also reduced binding. While mutants at the boundary of alpha helix (E23 and D28A) showed modest reduction (approximately 60% of wild-type binding) in their E6 binding ability, all other mutants made up of the alpha helical structure showed substantial decrease in E6 binding. Notably, a change of leucine to alanine at amino acid 25 totally abolished binding. The ability of mutants F18A and E23A to bind E6 confirmed the notion that the interaction of E6BP with E6 is independent of calcium binding, as both mutations abolished calcium binding. Finally, a leucine to proline change at amino acid 25 of E6bp was created and used in a binding experiment. This change (L25P) which is expected to disrupt the alpha helical structure, totally abolished E6bp-binding with E6, indicating that the alpha helix from the conserved motif is indeed important for E6 binding.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS:

1. A method for evaluating a candidate compound for the ability to interact with the HPV E6 transforming protein, comprising:

supplying a three-dimensional structure for E6bp;

supplying a three-dimensional structure for said candidate compound; and

comparing the three-dimensional structure of said candidate compound to the three-dimensional structure of said E6bp, wherein similarity in the structure of said candidate compound to the structure of said E6bp is indicative of the ability of said candidate compound to interact with said HPV E6 transforming protein,

thereby evaluating the ability of said candidate compound to interact with said HPV E6 transforming protein.

2. A method of providing or identifying a candidate compound which has the ability to interact with the HPV E6 transforming protein, comprising:

supplying a three-dimensional structure for E6bp;

supplying a three-dimensional structure for said candidate compound;

comparing the three-dimensional structure of said candidate compound to the three-dimensional structure of said E6bp; and

altering the structure or altering the spatial position of the structure, of said candidate compound,

thereby providing or identifying said candidate compound which has the ability to interact with said HPV E6 transforming protein.

3. The method of claim 2, further comprising comparing the altered structure or spatial position of the structure of said candidate compound to the three-dimensional structure of said E6bp.

4. The method of claim 3, further comprising making a second alteration in the structure or the spatial position of the structure, of said candidate compound.

5. The method of claim 2, wherein the method comprises defining an atom equivalency in said candidate compound and said E6bp three-dimensional structures.

6. The method of claim 2, wherein the method further comprises:

defining an atom equivalency in said candidate compound and said E6bp three-dimensional structures; and

performing a fitting operation between said candidate compound and said E6bp three-dimensional structures.

7. The method of claim 2, wherein said comparing step is performed visually.

8. The method of claim 2, wherein said alteration results in a change in the orientation of the structure of said candidate compound.

9. The method of claim 8, wherein said alteration comprises rotation of said structure.

10. The method of claim 9, wherein the orientation of the structure of said candidate compound is altered relative to the structure of the E6bp.

11. The method of claim 2, wherein the alteration results in a change in a parameter selected from the group consisting of: hydrophobicity, steric bulk, electrostatic properties, size, bond angle, type of bond and amino acid sequence of the candidate compound.

12. The method of claim 2, further comprising creating a record of one or both the three-dimensional structures of said candidate compound and said E6bp.

13. The method of claim 2, wherein the three-dimensional structure of said candidate compound and/or said E6bp is displayed on a machine capable of displaying a graphical three-dimensional representation.

14. The method of claim 2, wherein said candidate compound is selected from a database of three-dimensional structures.

15. The method of claim 14, wherein said three-dimensional structures are determined experimentally.

16. The method of claim 14, wherein said three-dimensional structures are computationally generated.

17. The method of claim 2, wherein said candidate compound is designed de novo.

18. A machine-readable data storage medium, comprising a data storage material encoded with a set of NMR derived coordinates which define the three-dimensional structure of the E6bp molecule.

19. A method of treating a subject at risk for infection by a HPV, comprising administering to a subject a therapeutically effective amount of a compound other than an E6bp fragment, wherein the candidate compound has a structure sufficiently duplicative of the structure of E6bp shown in Figure 1, so as to bind to the HPV E6 transforming protein and prevent its interaction with the ERC55 protein,

thereby treating a subject at risk for infection by a HPV.

20. A compound having a three-dimensional structure substantially similar to the three-dimensional structure of E6bp, such that the candidate compound can bind to the HPV E6 transforming protein, wherein said compound is other than E6bp or ERC55.

21. The compound of claim 20, wherein the structure of the candidate compound is sufficiently similar to the three-dimensional structure the E6bp provided in Figure 1, such that the candidate compound can bind to the HPV E6 transforming protein, with an affinity which is at least half that of E6bp.

22. A method of evaluating the ability of a candidate compound to associate with an E6bp molecule, comprising:

supplying a three-dimensional structure for said E6bp molecule;

supplying a three-dimensional structure for said candidate compound; and

performing a fitting operation between said three-dimensional structures of said candidate compound and said E6bp molecule,

thereby evaluating the ability of said candidate compound to associate with said E6bp molecule.

23. A method of modeling the region of the HPV E6 protein which binds ERC55, comprising:

supplying a three-dimensional structure for an E6bp molecule; and

supplying a complementary structure to said E6bp molecule,

thereby modeling the region of the HPV E6 protein which binds ERC55.

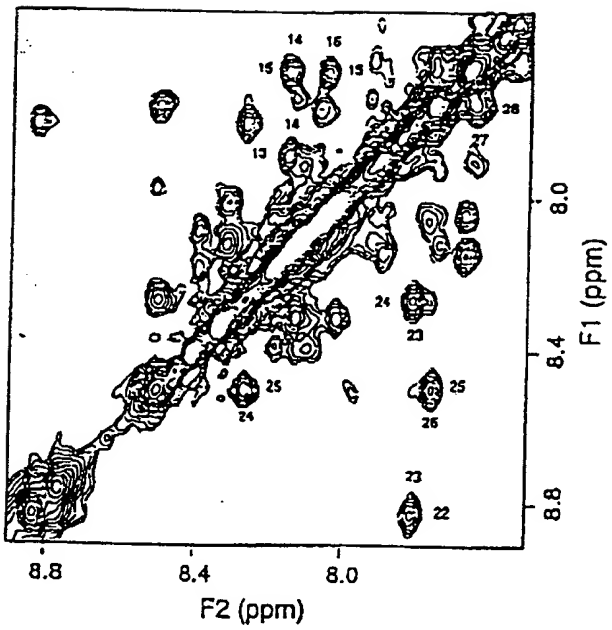

FIGURE 1

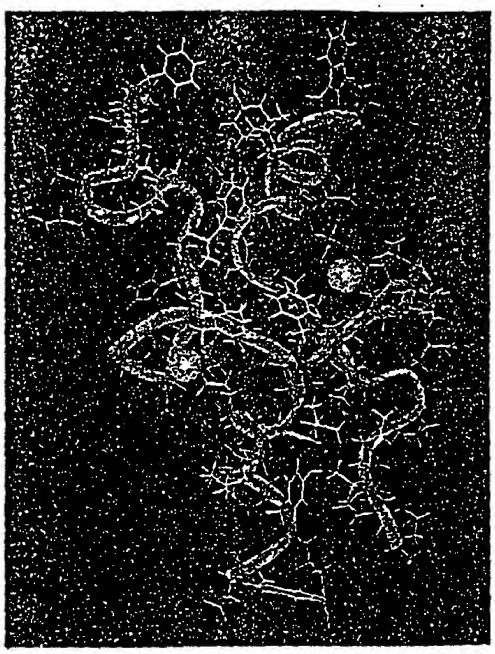

FIGURE 2

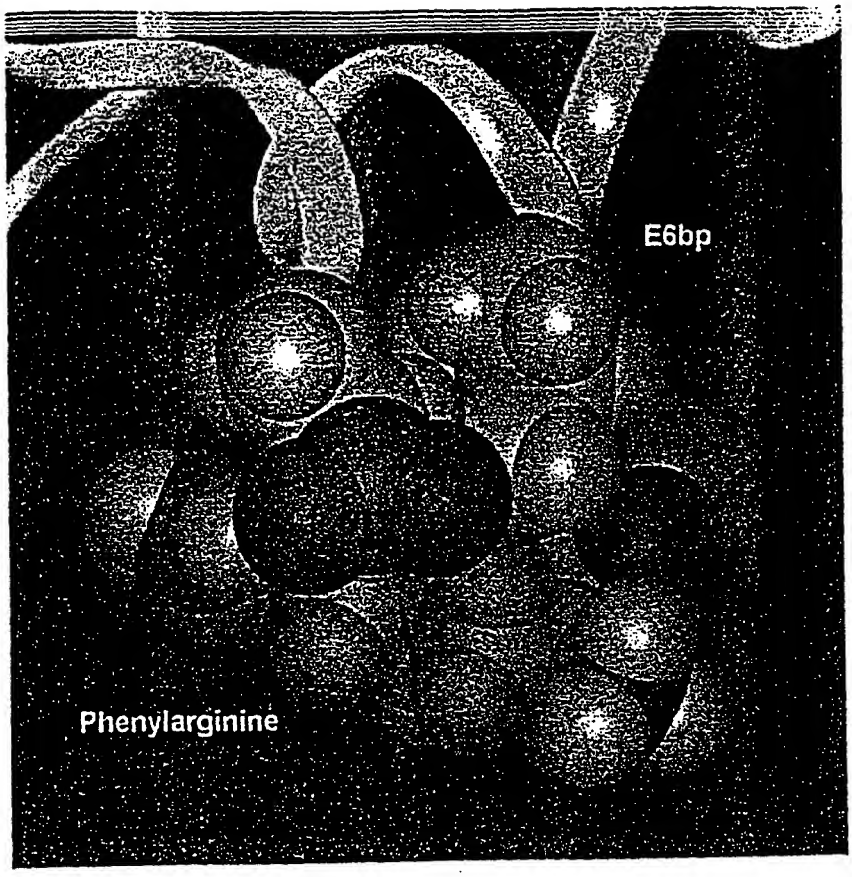

FIGURE 3

Identification f a short alpha helix from E6BP as E6 binding domain

SEQUENCES	Name	E6 BD
E L T L Q E L L G E E R M D D L D A L L A D L E S T L Q T L Q D I L G D P G D K	(E6-AP) (Paxillin LD1) (Tuberin)	
A L E E H D K N G D G F V S L E E F L G D Y R W D	WT	100%
D K N G D G F V S L E E F L G D Y R W D	dlH1	90%
D K N G D G F V S L E E F L G D Y R	dlH1C	52%
V S L E E F L G D Y R W D	dlH1L	39%
A L E E H D K N G D G **A** V S L E E F L G D Y R W D	F18A	35%
A L E E H D K N G D G F **A** S L E E F L G D Y R W D	V19A	141%
A L E E H D K N G D G F V **A** L E E F L G D Y R W D	S20A	114%
A L E E H D K N G D G F V S **A** E E F L G D Y R W D	L21A	30%
A L E E H D K N G D G F V S L **A** E F L G D Y R W D	E22A	155%
A L E E H D K N G D G F V S L E **A** F L G D Y R W D	E23A	65%
A L E E H D K N G D G F V S L E E **A** L G D Y R W D	F24A	6%
A L E E H D K N G D G F V S L E E F **A** G D Y R W D	L25A	0%
A L E E H D K N G D G F V S L E E F **P** G D Y R W D	L25P	0%
A L E E H D K N G D G F V S L E E F L **A** D Y R W D	G26A	26%
A L E E H D K N G D G F V S L E E F L G **A** Y R W D	D27A	18%
A L E E H D K N G D G F V S L E E F L G D **A** R W D	Y28A	57%
A L E E H D K N G D G F V S L E E F L G D Y **A** W D	R29A	133%
A L E E H D K N G D G F V S L E E F L G D Y R **A** D	W30A	71%

FIGURE 4

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: James D. Baleja, Elliot J.Androphy and Jason J.Chen

(ii) TITLE OF INVENTION:Structure-based Rational Design of Compounds to Inhibit Papillomavirus Infection

(iii) NUMBER OF SEQUENCES: 22

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: FISH & RICHARDSON
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02110

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/061,295
(B) FILING DATE: 07 October 1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Paul Louis Myers
(B) REGISTRATION NUMBER: 35,965
(C) REFERENCE/DOCKET NUMBER: NEP-006PC

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)542-5070
(B) TELEFAX: (617)542-8906

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Ile Pro Glu Ser Ser Glu Leu Thr Leu Gln Glu Leu Leu Gly Glu Glu
1 5 10 15

Arg Arg

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Glu Leu Thr Leu Gln Glu Leu Leu Gly Glu Glu Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asp Asp Leu Asp Ala Leu Leu Ala Asp Leu Glu Ser Thr
1 5 10

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu Gln Thr Leu Gln Asp Ile Leu Gly Asp Pro Gly Asp Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu Glu Phe Leu Gly Asp
1 5 10 15

Tyr Arg Trp Asp
20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu Glu Phe Leu Gly Asp
1 5 10 15

Tyr Arg

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Ser Leu Glu Glu Phe Leu Gly Asp Tyr Arg Trp Asp
1 5 10

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Ala Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Ala Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ala Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Ala Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Ala

1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Ala Phe Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Ala Leu Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Ala Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Pro Gly Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Ala Asp Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Ala Tyr Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Ala Arg Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Ala Trp Asp
20 25

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Ala Leu Glu Glu His Asp Lys Asn Gly Asp Gly Phe Val Ser Leu Glu
1 5 10 15

Glu Phe Leu Gly Asp Tyr Arg Ala Asp
20 25

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/20991

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A01N 61/00; A61K 45/00; C12Q 1/70; G06T 1/00
US CL : 435/5, 424/184.1+, 514/1+, 707/1+
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/5, 424/184.1+, 514/1+, 707/1+

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHEN et al. Interaction of Papillomavirus E6 Oncoproteins with a Putative Calcium-Binding Protein. Science. 28 July 1995, Volume 269, pages 529-531; see the document in its entirety and especially page 529, abstract and paragraphs 1-3.	1-23
Y	US 5,532,348 A (HUIBREGTSE ET AL.) 02 July 1996, see the abstract and column 1, line 1, through column 5, line 45.	1-23
Y	CHARRETIER et al. Application de la resonance magnetique nucleaire a la determination de la structure des proteines en solution. Biochemistry and Cell Biology. 1991, Volume 69, pages 323-335, see the English abstract beginning on page 322.	1-23

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 24 FEBRUARY 1999	Date of mailing of the international search report 12 MAR 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer BRENDA BRUMBACK Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/20991

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,583,000 A (ORTIZ DE MONTELLANO ET AL.) 10 December 1996, see column 1, lines 51-67; column 2, lines 36-39; column 8, lines 25-45; and column 35, lines 19-60.	1-23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/20991

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/20991

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, DIALOG: Medline, BIOTECH, Conf Papers, Euro, Japio, WPI
search terms: authors' names, papilloma, HPV, three-dimensional structure, nuclear magnetic resonance, NMR, analog(ue), treatment, therapeutic, pharmaceutical, E6

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1.

Group I, claims 1-17 and 19-22, drawn to compounds which bind HPV E6 transforming protein, method of making, and method of use.
Group II, claim 18, drawn to a data storage medium.
Group III, claim 23, drawn to a method of modeling the HPV E6 region which binds ERC-55.

The inventions listed as Groups I-III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of Group I relates to compounds which bind HPV E6 transforming protein. Groups II and III do not share this special technical feature. Therefore, they are not so linked as to form a single inventive concept under PCT Rule 13.1.